

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 10:47:49 ; Search time 2659.16 Seconds

(without alignments)
6923.559 Million cell updates/sec

Title: US-09-602-833A-1

Sequence: 1 atggagacataagtggtgtt.....ctttagccttcaacttga 1116

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

GenBank: 1: gb_ba: 2: gb_htg: 3: gb_in: 4: gb_cm: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_cm: 20: em_ov: 21: em_pat: 22: em_ph: 23: em_pl: 24: em_ro: 25: em_sts: 26: em_sy: 27: em_un: 28: em_vl: 29: em_htgo_hum: 30: em_htgo_inv: 31: em_htgo_inv: 32: em_htgo_inv: 33: em_htg_hum: 34: em_htg_inv: 35: em_htg_inv: 36: em_htg_inv: 37: em_htg_inv: 38: em_htg_inv: 39: em_htg_inv: 40: em_htg_inv: 41: em_htg_inv: 42: em_htg_inv: 43: em_htg_inv: 44: em_htg_inv: 45: em_htg_inv

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1116	100.0	1116	6	AX063232	AX063232 Sequence
2	1065	95.4	4860	9	HSA308569	AJ308569 Homo sapi
3	681	61.0	681	6	AX063234	AX063234 Sequence
4	209	18.7	145795	2	HS310996	AJ310996 Homo sapi
5	209	18.7	202324	2	AC068720	AC068720 Homo sapi
6	160	14.3	202324	2	AC068720	AC068720 Homo sapi
7	129	11.6	2056	9	AK021919	AK021919 Homo sapi
8	41	3.7	172966	9	AK013467	AK013467 Homo sapi
9	23	2.1	10213	3	PFU27338	U27338 Plasmid
10	23	2.1	115009	2	HS310996	HS310996 Homo sapi
11	22	2.0	155929	2	AC025574	AC025574 Homo sapi
12	21	1.9	65349	2	AL445472	AL445472 Homo sapi
13	21	1.9	77644	3	AC090521	AC090521 Caenorhab
14	21	1.9	109349	2	AC011381	AC011381 Homo sapi
15	21	1.9	146206	9	AC026800	AC026800 Homo sapi
16	21	1.9	163332	9	AC004821	AC004821 Homo sapi
17	21	1.9	175144	2	AC010415	AC010415 Homo sapi
18	21	1.9	180541	2	AC016521	AC016521 Homo sapi
19	21	1.9	192281	9	AC010424	AC010424 Homo sapi
20	21	1.9	193379	2	AC022182	AC022182 Homo sapi
21	21	1.8	217220	9	AC008682	AC008682 Homo sapi
22	21	1.8	410	1	RP282365	RP282365 R. prowazeki
23	20	1.8	573	4	CFAM100N	X98239 C. famillari
24	20	1.8	982	5	S7292353	S7292353 proenkephal
25	20	1.8	1076	5	XLENK02	X00853 xenopus lae
26	20	1.8	1359	5	XLENK01	X00852 xenopus lae
27	20	1.8	38749	3	AF099001	AF099001 Caenorhab
28	20	1.8	73174	2	AC036161	AC036161 Homo sapi
29	20	1.8	95937	8	AC007112	AC007112 Arabidops
30	20	1.8	103219	2	AC079745	AC079745 Homo sapi
31	20	1.8	120185	8	AC007138	AC007138 Arabidops
32	20	1.8	139480	2	AC012837	AC012837 Drosophila
33	20	1.8	146949	2	AL360003	AL360003 Homo sapi
34	20	1.8	147160	2	AC010047	AC010047 Drosophila
35	20	1.8	155941	9	HS310996	HS310996 Homo sapi
36	20	1.8	157959	2	AC021741	AC021741 Homo sapi
37	20	1.8	160264	9	AL442644	AL442644 Human DNA
38	20	1.8	161802	9	AL391819	AL391819 Human DNA
39	20	1.8	163790	2	AL590233	AL590233 Homo sapi
40	20	1.8	170610	2	AL161910	AL161910 Homo sapi
41	20	1.8	176174	9	AC007483	AC007483 Homo sapi
42	20	1.8	178429	2	AC026462	AC026462 Homo sapi
43	20	1.8	181360	2	AL356072	AL356072 Homo sapi
44	20	1.8	183457	2	AC092420	AC092420 Homo sapi
45	20	1.8	187093	2	AC079903	AC079903 Homo sapi

ALIGNMENTS

RESULT	1	AX063232	1116 bp	DNA	PAT	24-JAN-2001
LOCUS	AX063232	Sequence	1 from Patent WO0078959.			
DEFINITION	AX063232	Sequence	1 from Patent WO0078959.			
ACCESSION	AX063232	Sequence	1 from Patent WO0078959.			
VERSION	AX063232.1	GI:12541058				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						

Location/Qualifiers
1. 1116
/organism="Homo sapiens"
/db_xref="taxon:9606"

AEIGCKLNKELNVGNFYLNKSIPPELDCENLERLDCSGNLELMEJPELSNKKOYTF
 VDISANKFSSVPICVLMSNLMQDLIDSSNMLDPOIDLELSEOSPLKXNKLTLYP
 YSMNLKRLFLVYVSGDLVELPDLDDSSPTPKFVSIKMNPIIDNACEDGNEIMESE
 RDRQHPDKFEMKAYIEDLKRESEVSPSTTTVVSFSLQ"

BASE COUNT 1480 a 924 c 1162 g 1294 t
 ORIGIN

Query Match 95.4% Score 1065; DB 9; Length 4860;
 Best Local Similarity 99.9% Pred. No. 0;
 Matches 1115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggagacataaagtggtgtcttcgacattctgtcatcaagagcttgggaactcgt 60
 DB 168 ATGGAGACATAAAGTGTTGTTGACATTCTGTCAACACAGCCTTGTGGAAACTGT 227
 QY 61 gtcaagaagcaaaagcttggcagaagaagggtgaaaagcttgaagaagcgcttg 120
 DB 228 GTCAAGAAGCAAAAGCTTGGCAGAGAAGAGAGTGAAGAGTGAAGAGCGCTTG 287
 QY 121 gagaagataaaggagagtggaacttggcgcaatgcaaggaagagagcattcccccag 180
 DB 288 GAGAAGATTAAGAGAGAGTGGAACTTGTGGCCGAATGCAGAGAGGAAGGCAATCCCCAG 347
 QY 181 gctatatactgcaagaatggtctcatagacacacagcgctggcgcttcggagacagattga 240
 DB 348 GCTTATACCTGCAGAAATGGTTCAATGACACACAGCGGCTTGTGCAAGATTGCA 407
 QY 241 aggaacactcacaaggcagaagttcaactcccaagaagcagaagcagaagcagtgcg 300
 DB 408 AGGAACACTCTCACAAAGCAGAGTTCACCTCCCAAGGACAGCAAAAGGAGCAGTCCG 467
 QY 301 ttgtgtttaaacttctggggagacatgacgagctcccaagattatgaagaagcag 360
 DB 468 TTTGTGTTGAACCTTTCTGGGAGACCTGGAGGAGCTCCAGATTCATTGAAGAGCAG 527
 QY 361 acacacactgagaagaatggtcatataagaacatccttgattcaaacattccatataatt 420
 DB 528 ACACACTGAGAGAAATGGTACATAGCAATACCTGATTCAAATCCTACATATAT 587
 QY 421 caagtattcaagcgatgagaatctcgatctgcaaaaaaacaatccatattcca 480
 DB 588 CAGTATTTTCAAGAGATGAGAAATCTGATCTGCCAAAAAACCAAACTCCACATCTTCCA 647
 QY 481 gcaaaaactcggtgttggaaagaacctgaagaactaatggtgttcaaatatcgaag 540
 DB 648 GCAGAAATCGGTTTGTGAAGAACCTTAAGAACCTCAATGTGGGTTTCACTATCTGAG 707
 QY 541 agcatlccccaagaatggtgagattgtgaaatctagagagacgtgattgtctggaat 600
 DB 708 AGCATTCCTCCAGAAATGGGAGATTGTGAATAATCTAGAGACATGATTTGTGGAAT 767
 QY 601 ctgaatataatgagctgcgccttgaattgaattgaagaagttacattgtgat 660
 DB 768 CTAAATAATTAAGAGCTGCCCTTTGAATTAATTTGAAGCAAGTTACATTTGTGAT 827
 QY 661 atcccaagaagaagtttccagatgaccgaatctgttccggagatggtgaaatttcag 720
 DB 828 ATCTCACCAACAAGTTTCCAGTGTCCCAATCTGTGTCTCCGGAATGTCGAATTTGCA 887
 QY 721 tgggttgatatacgaagaacaaataaccgtgaccctgcccgaagatatagaagctagag 780
 DB 888 TGGTGTGATATACAGACATAATTAACCTGACCTGCCGGAAGATATAGACAGCTGAG 947
 QY 781 gaggatgaagagcttctctgtataaaaaaagaattgacataccttccattccatgctg 840
 DB 948 GAGGTGAAGAGCTTCTCTGTATTAATAAACAAAGTTGACATCACTTCCATTTCCATCTG 1007
 QY 841 aacctgaagaagctcactcgttagtctgcaatgggagacattggggagagctcccaact 900
 DB 1008 AACTGGAAGAAGCTCACTCTGTAGTGTGCTGAGTGGGAGCACTTTGGTGGAGACTCCCACT 1067
 QY 901 gcccttgtgactatccacacacttaaaatttgaagccttatggaacaatccattgat 960

DB 1068 GCCCTTGTGACTATCCACACCTTTAAATTTGTAAGCCTTATGAGCAATCCATATGAT 1127
 QY 961 aatgcccaatgtaagaatggaatgaatgaatgaagaagtgagacggagcccaacttt 1020
 DB 1128 AATGCCCAATGTGAAGAGGGAATGAATAATGAAAGTGAACGGAGATCCCAACATTTT 1187
 QY 1021 gataaagagttatgaagaagcctatatitgaagaccttaagaagaagaatcgttccagc 1080
 DB 1188 GATTAAGAAGTTATGAAGAGCCTATATTTGAAGACCTTAAGAAGAGAAATCTGTCCAGC 1247
 QY 1081 tataccaccaagtgctctttagccttcaacttga 1116
 DB 1248 TATACACCAAGTGTCTTTAGCCTTCAACTTTGA 1283

RESULT 3
 AX063234
 LOCUS AX063234 681 bp DNA PAT 24-JAN-2001
 DEFINITION Sequence 3 from Patent WO0078959.
 ACCESSION AX063234
 VERSION AX063234.1 GI:12541060
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 681)
 TURNER,A.C., Zamborwitz,B., Nehls,M., Friedlich,G.A. and Sands,A.T.
 Human genes and proteins encoded thereby
 Patent: WO 0078959-A 3 28-DEC-2000;
 JOURNAL Lexicon Genetics Incorporated (US)

FEATURES
 source Location/Qualifiers
 1..681
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..681
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC25078.1"
 /db_xref="GI:12541061"
 /translation="MRILDPKNOISHPAELICLNKELNVGNFYLNKSIPPELDC
 ENLERLDCSGNLELMEJPELSNKKOYTFVDSANKFSSVPICVLMSNLMQDLIDSSN
 NLTDLPDIDRLELOSLFLYKXKRLTYPLYSMLNKKLTLLVSGDLVELPDLDCS
 SPTLKFVSIKMNPIIDNACEDGNEIMESEDRQHPDKFEMKAYIEDLKRESEVSPSTT
 KVSFSLQ"

CDS

BASE COUNT 212 a 138 c 142 g 189 t
 ORIGIN

Query Match 61.0% Score 681; DB 6; Length 681;
 Best Local Similarity 100.0% Pred. No. 0;
 Matches 681; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 atggaatttcgagatcgcgcaaaaacaacatccacatcttcagaagaatcggtgt 495
 DB 1 ATGGAATTCGAGATCTGCCAAAAACCAATCTCACATCTTCCAGCAGAAATGGTTGT 60
 QY 496 ttgaagaacctgaagaactcaatggtgttcaactatctgaagaagcattcccgaa 555
 DB 61 TTGAAGAACCCTGAAGAACCTCAATGTGGGTTTCAACTATCTGAAGACATCTCCAGAA 120
 QY 556 ttgggaagatttgaaaatctagaagaagctgattgttctggaatctagaatgaatgag 615
 DB 121 TTGGAGATTGTGAATACTAGAGAGACTGATTGTTCTGAAAATCTAGAAATTAATGAG 180
 QY 616 ctgcgccttgaatgaatlaatttgaagcaagttacattgtatagatctcagaacaaga 675
 DB 181 CTGCCCTTTGAATTAAGTAATTTGACAGAGTTTCAATTTGTAGATATCTCAGCAAAAG 240
 QY 676 ttctcagatgtcccaatctgttctcgcgagatgctgaatttgcagtggttgatatagc 735
 DB 241 TTTTCCAGTGTCCCAATCTGTGTCTGCGATGTGGAATTTGCAAGTGTGGATATCAGC 300

```

Oy 756 agcaataactgaccgacctgcgcgaagatataagacaggtagagagctgcagagctt 795
Db 301 AGCAATAAACCCTGACCGACCTGCCGCAAGATATAGACAGCTAGAGAGCTGCAGAGCTT 360
Oy 756 cctctgtataaaacaagttgacctactctccctattccatgctgtaacctgaagaagctc 855
Db 351 CTTCTGTATATAAACAAGTGAAGTGAACCTCCCTATTCATCTGTAACCTGAAGAAGCTC 420
Oy 856 actctgtatgctgtagtgaggacattgtgtgagagctcccaactgacctgtgtactca 915
Db 421 ACTCTGTATAGTGTGCTAGTGGGACCATTTGGTGGAGCTCCCAACTGCTTTGTGACTCA 480
Oy 916 tccacaccttaaaattgtgaagccttatgacaaatccctattgataatgcccacatgtga 975
Db 481 TCCACACCTTTAAATTGTSTAAGCCTTATGACAAATCTTATGATATATGCCCAATGTGAA 540
Oy 976 gatgcaatgaataataatggaagaatggaagcgatcgccaaacatttggataagaagtattg 1035
Db 541 GATGCAATGAATAATATGGAAGTGAACGCGATCCCAACATTTGTATGAAGAGTTATG 600
Oy 1036 aaagcctatattgaagaccttaagaagaagatctgttcccgacctatacccaaatg 1095
Db 601 AAAGCCTATATTGAAGACCTTTAAGAAGAAGATCTGTCCACGCTATATCCACCAAGTGC 660
Oy 1096 tctttaagcctcaacttga 1116
Db 661 TCTTTAGCCTTCACACTTGA 681

RESULT 4
HSA310996/c HSA310996 145795 bp DNA HTG 18-APR-2001
LOCUS Homo sapiens chromosome 3 clone RP6-91P17 map 3p21.3, ***
DEFINITION SEQUENCING IN PROGRESS ***, 42 ordered pieces.
ACCESSION AJ310996
VERSION AJ310996.1 GI:13751170
KEYWORDS HTG, HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 145795)
AUTHORS Kiss,H., Yang,Y., Kost-Alimova,M., Seles,A., Kholodnyuk,I.,
Kedra,D., Kiss,C., Klein,G., Imreh,S. and Dumanski,J.P.
Transcriptional map of the common eliminated region 1 on human
3p21.3
TITLE Unpublished
JOURNAL 2 (bases 1 to 145795)
REFERENCE Kiss,H.
AUTHORS Direct Submission
JOURNAL Submitted (17-APR-2001) Kiss H., Microbiology and Tumorbiology
Center (MTC), Karolinska Institute, Box 280, Stockholm, S-17177,
SWEDEN
COMMENT The sequencing contigs are in order and the gaps between them are
represented by 100 n.s.
Contig 1: 1-2508 Contig 2: 2609-17434 Contig 3: 17535-21987
Contig 4: 22088-23066 Contig 5: 23167-28845
Contig 6: 28946-30840 Contig 7: 30941-37879 Contig 8:
37980-38928 Contig 9: 39029-40834 Contig 10: 40935-44527 Contig
11: 44628-48077 Contig 12: 48178-52755 Contig 13: 52856-53293
Contig 14: 53394-54768 Contig 15: 54869-57895 Contig 16:
57966-59322 Contig 17: 59423-60159 Contig 18: 60260-61231
Contig 19: 61332-62928 Contig 20: 63029-63852 Contig 21:
63935-75555 Contig 22: 75656-77760 Contig 23: 77861-79686
Contig 24: 79787-80857 Contig 25: 80958-87622 Contig 26:
87723-89164 Contig 27: 89265-90077 Contig 28: 90108-92035
Contig 29: 92136-97298 Contig 30: 97389-102032 Contig 31:
102133-106605 Contig 32: 106706-108263 Contig 33: 108364-110022
Contig 34: 110133-113747 Contig 35: 113848-125151 Contig 36:
125257-126026 Contig 37: 126127-129649 Contig 38: 129750-132160
Contig 39: 132261-133125 Contig 40: 133226-135661 Contig 41:
135762-142148 Contig 42: 142249-145795.

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2508: contig of 2508 bp in length
* 2609 2608: gap of 100 bp
* 17435 17534: contig of 14826 bp in length
* 17535 21987: contig of 4453 bp in length
* 21988 22087: gap of 100 bp
* 22088 23066: contig of 979 bp in length
* 23067 23166: gap of 100 bp
* 23167 28845: contig of 5679 bp in length
* 28846 28945: gap of 100 bp
* 28946 30840: contig of 1895 bp in length
* 30841 30940: gap of 100 bp
* 30941 37879: contig of 6939 bp in length
* 37880 37979: gap of 100 bp
* 37980 38928: contig of 949 bp in length
* 38929 39028: gap of 100 bp
* 39029 40834: contig of 1806 bp in length
* 40835 40934: gap of 100 bp
* 40935 44527: contig of 3593 bp in length
* 44528 44627: gap of 100 bp
* 44628 48077: contig of 3450 bp in length
* 48078 48177: gap of 100 bp
* 48178 52755: contig of 4578 bp in length
* 52756 52855: gap of 100 bp
* 52856 53293: contig of 438 bp in length
* 53294 53393: gap of 100 bp
* 53394 54768: contig of 1375 bp in length
* 54769 54868: gap of 100 bp
* 54869 57895: contig of 3027 bp in length
* 57896 57995: gap of 100 bp
* 57996 59322: contig of 1327 bp in length
* 59323 59422: gap of 100 bp
* 59423 60159: contig of 737 bp in length
* 60160 60259: gap of 100 bp
* 60260 61231: contig of 972 bp in length
* 61232 61331: gap of 100 bp
* 61332 62928: contig of 1597 bp in length
* 62929 63028: gap of 100 bp
* 63029 63852: contig of 824 bp in length
* 63853 63952: gap of 100 bp
* 63953 75555: contig of 11603 bp in length
* 75556 75655: gap of 100 bp
* 75656 77760: contig of 2105 bp in length
* 77761 77860: gap of 100 bp
* 77861 79686: contig of 1826 bp in length
* 79687 79786: gap of 100 bp
* 79787 80857: contig of 1071 bp in length
* 80858 80957: gap of 100 bp
* 80958 87622: contig of 6665 bp in length
* 87623 87722: gap of 100 bp
* 87723 89164: contig of 1442 bp in length
* 89165 89264: gap of 100 bp
* 89265 90007: contig of 743 bp in length
* 90008 90107: gap of 100 bp
* 90108 92035: contig of 1928 bp in length
* 92036 92135: gap of 100 bp
* 92136 97298: contig of 5163 bp in length
* 97299 97398: gap of 100 bp
* 97399 102032: contig of 4634 bp in length
* 102033 102132: gap of 100 bp
* 102133 106605: contig of 4473 bp in length
* 106606 106705: gap of 100 bp
* 106706 108263: contig of 1558 bp in length
* 108264 108363: gap of 100 bp

```

```
* 108364 110022: contig of 1659 bp in length
* 110023 110122: gap of 100 bp
* 110123 113747: contig of 3625 bp in length
* 113748 113847: gap of 100 bp
* 113848 125151: contig of 11304 bp in length
* 125152 125251: gap of 100 bp
* 125252 126026: contig of 775 bp in length
* 126027 126126: gap of 100 bp
* 126127 129649: contig of 3523 bp in length
* 129650 129749: gap of 100 bp
* 129750 132160: contig of 2411 bp in length
* 132161 132260: gap of 100 bp
* 132261 133125: contig of 865 bp in length
* 133126 133225: gap of 100 bp
* 133226 135661: contig of 2436 bp in length
* 135662 135761: gap of 100 bp
* 135762 142148: contig of 6387 bp in length
* 142149 142248: gap of 100 bp
* 142249 145795: contig of 3547 bp in length.
Location/Qualifiers
1.145795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21.3"
/clone="RP6-91p17"
BASE COUNT 39764 a 33207 c 31499 g 37225 t 4100 others
ORIGIN

Query Match 18.7%; Score 209; DB 2; Length 145795;
Best Local Similarity 100.0%; Pred. No. 6.2e-107;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 gataaaggagagtggaacttgctgagcgaatcagaagaagggcaatccccaagctgt 185
|||||
DB 73652 GATTAAGGAGAGAGTGGAACTTGTGGCGAATCGACGAGAGAGGCAATCCCGACGCTGT 73593
|||||
QY 186 atactcaagaatgagctctacacccagcagctgagcgtcttgacaaagttaaagaa 245
|||||
DB 73592 ATACTGCAGAAATGCTTCTTAGACACACAGCGTGCGCTTCTGACAGAAATGAAAGAA 73533
|||||
QY 246 cacttcacaaagcagaagtcacttcccaagacagaagcaacagagacagtcgattgt 305
|||||
DB 73532 CACTTCACAAAGCAGAGTTCACATCCCAAGAGAGAGCAAGGAGAGAGTGGCTTTGT 73473
|||||
QY 306 gtttgaacttctggggagacagtgacg 334
|||||
DB 73472 GTTTGAACCTTCTGGGAGACAGTGCAGCG 73444
|||||

RESULT 5
AC068720/c 202324 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 3 clone RP11-509121, WORKING DRAFT
DEFINITION
AC068720
ACCESSION AC068720.2 GI:8469022
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 202324)
REFERENCE 1
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 202324)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
```

```
COMMENT
On Jun 12, 2000 this sequence version replaced gi:7715661.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0509121
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 193150 bases at least Q40
Consensus quality: 195686 bases at least Q30
Consensus quality: 197015 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 200424; sum-of-contigs
Quality coverage: 4.67 in Q20 bases; agarose-fp
Quality coverage: 4.72 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1576: contig of 1576 bp in length
1577 1676: gap of unknown length
1677 3263: contig of 1587 bp in length
3264 3363: gap of unknown length
3364 5443: contig of 2080 bp in length
5444 5544: gap of unknown length
5544 8196: contig of 2653 bp in length
8197 8296: gap of unknown length
8297 12038: contig of 3742 bp in length
12039 12138: gap of unknown length
12139 15992: contig of 3854 bp in length
15993 16092: gap of unknown length
16093 21192: contig of 5100 bp in length
21193 21292: gap of unknown length
21293 24766: contig of 3474 bp in length
24767 24866: gap of unknown length
24867 30824: contig of 5958 bp in length
30825 30924: gap of unknown length
30925 37153: contig of 6229 bp in length
37154 37253: gap of unknown length
37254 45262: contig of 8009 bp in length
45263 45362: gap of unknown length
45363 56733: contig of 11371 bp in length
56734 56833: gap of unknown length
56834 72156: contig of 15333 bp in length
72157 72256: gap of unknown length
72257 84702: contig of 12446 bp in length
84703 84802: gap of unknown length
84803 101052: contig of 16250 bp in length
101053 101152: gap of unknown length
101153 117709: contig of 16557 bp in length
117710 117809: gap of unknown length
117810 132650: contig of 14481 bp in length
132651 132750: gap of unknown length
132751 152071: contig of 19321 bp in length
152072 152171: gap of unknown length
152172 176190: contig of 24019 bp in length
176191 176290: gap of unknown length
176291 202324: contig of 26034 bp in length.
Location/Qualifiers
1.202324
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
/chromosome="3"
/clone="RP11-509121"
1.1576
misc_feature
/note="assembly_name:Contig4"
1677..3263
misc_feature
/note="assembly_name:Contig5"
3364..5443
misc_feature
/note="assembly_name:Contig6"
5544..8196
misc_feature
/note="assembly_name:Contig7"
8297..12038
misc_feature
/note="assembly_name:Contig8"
12139..15992
misc_feature
/note="assembly_name:Contig9
clone_end:T7
vector_side:right"
16093..21192
misc_feature
/note="assembly_name:Contig10"
21293..24766
misc_feature
/note="assembly_name:Contig11"
24867..30824
misc_feature
/note="assembly_name:Contig12"
30925..37153
misc_feature
/note="assembly_name:Contig13"
37254..45262
misc_feature
/note="assembly_name:Contig14"
45363..56733
misc_feature
/note="assembly_name:Contig15"
56834..72156
misc_feature
/note="assembly_name:Contig16"
72257..84702
misc_feature
/note="assembly_name:Contig17"
84803..101052
misc_feature
/note="assembly_name:Contig18"
101153..117709
misc_feature
/note="assembly_name:Contig19"
117810..132650
misc_feature
/note="assembly_name:Contig20
clone_end:SP6
vector_side:left"
132751..152071
misc_feature
/note="assembly_name:Contig21"
152172..176190
misc_feature
/note="assembly_name:Contig22"
176291..202324
misc_feature
/note="assembly_name:Contig23"

BASE COUNT 56787 a 45300 c 43765 g 54528 t 1944 others
ORIGIN
```

```
Query Match 18.7% Score 209; DB 2; Length 202324;
Best Local Similarity 100.0%; Pred. No. 6.4e-107;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 gataaaggagagtggaacttgcgcgaatgcaggaaggacatcccccagcgtcgt 185
|||||
DB 30421 GATTAAGAGAGAGTGGAAGCTTTGCGCGAATGACAGAGAGGCGATCCCGAGGCGTGT 30362
|||||
QY 186 ataccgcaagaatgcttcacagacacacagcgtcgcgcttcgtgacagaattgaaagaa 245
|||||
DB 30361 ATACTGCAAGAAATGCTTCATAGACACACGCTGCGGCTTGTGACAAAGATTGAAGAA 30302
|||||
QY 246 cacttcacaagcagaagttcactcccaagaagacagagcaaacgagcagtcgcttgc 305
|||||
DB 30301 CACTCTCACAAGGACGAGACTTCCCAAGGACAGAGCAAGGAGACGACGAGTGGCTTGT 30242
|||||
QY 306 gtttgaacttctcgggagacactggaagg 334
|||||
DB 30241 GTTTGAACCTTCTGGGGAGCAGTGCACG 30213
|||||
```

RESULT 6
AC068720

```
LOCUS AC068720 202324 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-509121, WORKING DRAFT
ACCESSION AC068720
VERSION AC068720.2 GI:8469022
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 202324)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 202324)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 12, 2000 this sequence version replaced gi:7715661.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information -----
Center project name: H.NH0509121
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 193150 bases at least Q40
Consensus quality: 197015 bases at least Q30
Insert size: 200000; agarose-fp
Insert size: 200424; sum-of-contigs
Quality coverage: 4.67 in Q20 bases; agarose-fp
Quality coverage: 4.72 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1576: contig of 1576 bp in length
* 1577 1676: gap of unknown length
* 1677 3263: contig of 1587 bp in length
* 3263 3364: gap of unknown length
* 3364 5443: contig of 2080 bp in length
* 5443 5544: gap of unknown length
* 5544 8196: contig of 2653 bp in length
* 8196 8297: gap of unknown length
* 8297 12038: contig of 3742 bp in length
* 12038 12139: gap of unknown length
* 12139 15992: contig of 3854 bp in length
* 15992 16093: gap of unknown length
* 16093 21192: contig of 5100 bp in length
* 21192 21293: gap of unknown length
* 21293 24766: contig of 3474 bp in length
* 24766 24867: gap of unknown length
* 24867 30824: contig of 5958 bp in length
* 30824 30925: gap of unknown length
* 30925 37153: contig of 6229 bp in length
* 37153 37254: gap of unknown length
* 37254 45262: contig of 8009 bp in length
* 45262 45363: gap of unknown length
* 45363 56733: contig of 11371 bp in length
```

```

* 56734 56833: gap of unknown length
* 56834 72156: contig of 15323 bp in length
* 72157 72256: gap of unknown length
* 72257 84702: contig of 12446 bp in length
* 84703 84802: gap of unknown length
* 84803 101052: contig of 16250 bp in length
* 101053 117709: contig of 16557 bp in length
* 117710 117810: gap of unknown length
* 117810 132650: contig of 14841 bp in length
* 132651 132750: gap of unknown length
* 132751 152071: contig of 19321 bp in length
* 152072 176190: contig of 24019 bp in length
* 176191 176291: gap of unknown length
* 176291 202324: contig of 26034 bp in length.

```

FEATURES

```

source

```

```

1. .202324
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-509I21"
1. .1576
/feature="assembly_name:Contig4"
1677. .3263
/feature="assembly_name:Contig5"
3364. .5443
/feature="assembly_name:Contig6"
5544. .8196
/feature="assembly_name:Contig7"
8297. .12038
/feature="assembly_name:Contig8"
12139. .15992
/feature="assembly_name:Contig9"
clone_end:T7
vector_side:right"
16093. .21192
/feature="assembly_name:Contig10"
21293. .24766
/feature="assembly_name:Contig11"
24867. .30824
/feature="assembly_name:Contig12"
30925. .37153
/feature="assembly_name:Contig13"
37254. .45262
/feature="assembly_name:Contig14"
45363. .56733
/feature="assembly_name:Contig15"
56834. .72156
/feature="assembly_name:Contig16"
72257. .84702
/feature="assembly_name:Contig17"
84803. .101052
/feature="assembly_name:Contig18"
101153. .117709
/feature="assembly_name:Contig19"
117810. .132650
/feature="assembly_name:Contig20"
clone_end:SP6
vector_side:left"
132751. .152071
/feature="assembly_name:Contig21"
152172. .176190
/feature="assembly_name:Contig22"
176291. .202324
/feature="assembly_name:Contig23"

```

```

BASE COUNT 56787 a 45300 c 43765 g 54528 t 1944 others
ORIGIN

```

```

Query Match 14.3%; Score 160; DB 2; Length 202324;
Best local Similarity 100.0%; Pred. No. 4.7e-79;
Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 333 ggaagccagattcattgaagagagacacacccctagagaatggtacataaacaatc 392
|||||
Db 85258 GGAGCTCCGAGTTCATTGGAAGAGAGACACACCTGAGAGATGATGACATTAACCAATAC 85317
Oy 393 ctgatcacaatcattccatataatcattatcattcaagcagatgagaattcgatct 452
|||||
Db 85318 CTGATTCACATATCTCTCATATATTCAGTATTTTCAAGCGATGAGAAATTCGATCT 85377
Oy 453 gccaaaaaaccaatcctcaatcttcacagcagaatcggt 492
|||||
Db 85378 GCCAAAAAACCAATCTCATATCTCCAGCAGAAATCGGT 85417

```

RESULT 7

```

AK021919
LOCUS AK021919 2056 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA FLJ11857 fis, clone HEMBA1006807, moderately
similar to Homo sapiens mRNA for SPOF.
ACCESSION AK021919
VERSION AK021919.1 GI:10433216
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
mRNA, clone_11b:HEMBA1006807.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugeno,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiyasu,S., Komai,F., Hara,R., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Makamatsu,A., Nakamura,Y., Nagahara,K., Masuno,Y. and Oshima,A.
MEDO human cDNA sequencing project
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugeno,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiyasu,S., Komai,F., Hara,R., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Makamatsu,A., Nakamura,Y., Nagahara,K., Masuno,Y. and Oshima,A.
MEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2056)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)

```

```

COMMENT
MEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan, cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

```

FEATURES

```

source
1. .2056
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1006807"
/clone_11b="HEMBA1"
/seq_stage="embryo, 10 weeks"
/tissue_type="whole embryo, mainly head"
/feature="cloning vector: pME185FL3"
247. .1422
/feature="unnamed protein product"
/codon_start=1
/protein_id="BAH13937.1"
/db_xref="GI:10433217"
/translation="MSREPPPLPGMSGTPPIAESNCYQVYKYSMTTINSEFC
REMEGVLAKESTFSSGSDMKWCLVNPGLDDESKDYLSTLLVSCPSVRAKE
KESILNAKRETKAMESORAYRVOGKMGFKFIRRDPLDEANLIDPDKSLTJLCE
SVSVQDSVNSIGHTNNTLKREPCRLAEDIGNMENTRFDGCFVARGOEFKAKSVL
AARSPVNAFMEHEMESKKNREIINDLDPEVKEEMRFTYCGRAVNDLKMADNLLA
ADKYALERLKYGEALCSLSYENVADTIVLADLHSAQDKRAQADIPNRCVLEQL
GCKDGNMNSNQATDIMESEKRDQHFDEKEMKAVIEDLKERESVPSYITTKVSPFDL"

```

CDS

```

BASE COUNT 642 a 394 c 495 g 525 t
ORIGIN

```

Query Match	11.6%	Score 129	DB 9	Length 2056
Best Local Similarity	100.0%	Pred. No. 1.5e-61		
Matches 129	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy	988	ataatcgaagtcgacggatcgcacaactttgtataaagaagttagaagcctataatt	1047	
Db	1294	ATTAATGGAAGATGAACGGGATGCCAACATTTGTGATTAAGAGATTATGAAGGCTTATT	1353	
Oy	1048	gaagacctaagaagaagaaatcgttctccgaactataccaccaagaatgtcctttagcctt	1107	
Db	1354	GAAAGACCTTAAAGAAAGAGAATCTGTCTCCACCTATACCAACCAAACTGCTTTAGCCTT	1413	
Oy	1108	caacttga	1116	
Db	1414	CAACTTGA	1422	
RESULT	8			
AC013467/c				
LOCUS	AC013467	172966 bp	DNA	PRI 25-MAY-2001
DEFINITION	Homo sapiens clone RP11-451F14, complete sequence.			
ACCESSION	AC013467			
VERSION	AC013467.8	GI:14196420		
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 172966)			
JOURNAL	Waterston, R.H.			
REFERENCE	The sequence of Homo sapiens clone			
AUTHORS	2 (bases 1 to 172966)			
TITLE	Waterston, R.H.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (12-NOV-1999) Genome Sequencing Center, Washington			
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
TITLE	MO 63108, USA			
JOURNAL	3 (bases 1 to 172966)			
REFERENCE	Waterston, R.H.			
AUTHORS	Direct Submission			
TITLE	Submitted (25-MAY-2001) Genome Sequencing Center, Washington			
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
REFERENCE	MO 63108, USA			
AUTHORS	On May 25, 2001 this sequence version replaced gi:13624417.			
TITLE	Center project name: H.NH051F14.			
COMMENT				
FEATURES				
Source	Location/Qualifiers			
	1..172966			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="RP11-451F14"			
BASE COUNT	57371 a 34327 c 31782 g 49486 t			
ORIGIN				
Query Match	3.7%	Score 41	DB 9	Length 172966
Best Local Similarity	100.0%	Pred. No. 2.2e-11		
Matches 41	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy	248	ctctcacaagcagagcttactctccccaagagacagagcaaa	288	
Db	169114	CTCTCACAAGCGAGAGTTCATCTCCCAAGGACAGAGGCMAA	169074	
RESULT	9			
PF027338				
LOCUS	PF027338	10213 bp	DNA	INV 27-FEB-1996
DEFINITION	Plasmodium falciparum erythrocyte membrane protein 1, Malayan Camp			
ACCESSION	U27338			
VERSION	U27338.1	GI:914918		
KEYWORDS				

[illegible]


```

CINGCNKICBVCARWLEIKNGENGNKIKKHYNINSNDKETIAVNVKSYFVDOGLFDID
YKRAQKVEDERKIMCTGHDCESEKEREKNEITNLISELOKITSCONKHP
NGKTAQDPSPPTPEEDPLDDTPOPLDDDOHBOKPCPPPMPTCYEKIAXELR
VEAEKINNELKNGKDGKCNVKKKNGAVIGESCEFEQYENSNNINKCND
ONERFKIGOKMFKYIGTIRKDLICIPRREHMCIDLISLGRITTSLSALKIDBA
AKSERDIDIRKLEQNSCDEHRICDARKYFADLGDILRGDLMNKNKADAKELKRP
YAFINITYNLQNDKNKYEKDRPYLQIRSDMOPANRHHINAMTNNKAPDARKLKP
NDTSSSSSKGIMTTHNSGIDKPEPDYDIPOFRMOEMSEFCILNEEMEOFEK
TGCECKNSTITCEDDRGTGKENCNCKCEKYLHNMKIGPDKYKELINEIYNNDS
KINSNEKFKELKDKCKELNSDKCIDEAHCITKYSKNSENKNNHNNYAKRNP
EYKACACDAPLDDNCPDSATYERKACNLPLPKLCESTFENNDDSWOTSFQTS
RDNTGVLVPPRRQICLKNTITRLSLEKIDDRKAEIMTSAYNEGKLCILYKRDY
TLOAMKYSFPRDQIDYKDLISTAPLSTKLNLVLKGDGNEIKEDGKMTYER
TRVHMLGKCYKAAGKIEERDCSLPDNTHOFLRMFREMSEHCAROKLPEVKEE
CASACCIIEYGTIDPPYCEACQOYRIYIRKIOEYRLWYOYNTNNEKAEYTA
EYFNDCNDKCNCLSKYIDIEKKMKNNYDSEPDNDLKNKICQIRKRPKRYKEE
EHTPSBODTPPLPPKDDLPPEEPFNDILKTIPEFIALALGSLATLFLKTK
SSVGNLFQILHPRSDYDIPKLSPNRYIYTSKYGKRYIYLEGSDGSDGYTHY
SDITSSSESEYELINDIYVPGSPKYLEVLEPSSGNNVTASGKNTPSDONDIO
NDGIPSKITDNEMNTLKDEFISMLONENEPNMLGYVNDNHTPTSRHNYEKP
FMSIHDRDLYSGEESYNNVNNNDIPISARNGNSGIDLINDSLNSKVVYIDEL
KRKENLFCGNHMKNNSTNSVAKNTPTDINHOMLFHMLDRHRMCKEPTNNK
EELLDKLKEKMNKONNSGNTNPSGNTPIPSDIPSGKOSDIPSDNINHSDIPLYLNDV
SIQIHMDNPPINEFSMDYPPNNSMDTILEDLKPFNPPYDYVDQDDIYDVHDD
TSTVDNAMDPSKVQIEMDVNTKLYEKYPIADLMDI"
555
/variation
/clone="MCvar-1 pfEMP1"
/clone="A2 of MC-pfEMP1"
/replace="C"
557
/variation
/clone="MCvar-1 pfEMP1"
/clone="A2 of MC-pfEMP1"
/replace="A"
660
/variation
/clone="MCvar-1 pfEMP1"
/clone="A2 of MC-pfEMP1"
/replace="G"
1328
/variation
/clone="MCvar-1 pfEMP1"
/clone="B2 of MC-pfEMP1"
/replace="G"
1753
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="T"
1965
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="A"
2021
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="G"
2094
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="A"
2136
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="G"
2218
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="A"
2410
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="T"
2602
/variation
/clone="MCvar-1 pfEMP1"
/clone="C2 of MC-pfEMP1"
/replace="A"

```

```

Intron
7673. .8397
/gene="MCvar-1 pfEMP1"
exon
8398. .10213
/number=2
3'UTR
9745. .10213
BASE COUNT
4170 a 1429 c 1869 g 2745 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 118 ttggagaagaagaagagtg 140
Db 3854 TTGGAGAGATTAAGAGAGAGTG 3876
RESULT 10
HSDJ92C4
LOCUS
DEFINITION
Human DNA sequence from clone RP1-92C4 on chromosome 6q14.1-15
Contains ESTs, STS and GSSs. Contains the 3' part of a novel gene,
complete sequence.
ACCESSION
AL132875
VERSION
AL132875.22 GI:10334642
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 115009)
REFERENCE
Dunn, M.
Direct Submission
Submitted (21-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 27, 2000 this sequence version replaced gi:10190642.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
RP1-92C4 is from the library RP1-1 constructed by the group of
Pleiter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RP1-92C4
it may be shorter because we sequence overlapping sections only
once, except for a 100 base overlap.
The true left end of clone RP1-92C4 is at 1 in this sequence. The
true left end of clone RP3-357D13 is at 114910 in this sequence.
FEATURES
source
1. 115009
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q14.1-15"

```

```
repeat_region /clone="RP1-92C4"
                /clone_lib="RPC1-1"
                26. .323
repeat_region /note="AluJo repeat: matches 2. .289 of consensus"
                339. .489
repeat_region /note="L2 repeat: matches 2033. .2171 of consensus"
                490. .798
repeat_region /note="AluJb8 repeat: matches 1. .299 of consensus"
                799. .1242
repeat_region /note="L2 repeat: matches 1532. .2033 of consensus"
                1636. .1794
repeat_region /note="LTR45 repeat: matches 366. .525 of consensus"
                1868. .1947
repeat_region /note="LTR45 repeat: matches 52. .133 of consensus"
                2383. .3011
repeat_region /note="MER82 repeat: matches 1. .653 of consensus"
                3150. .3197
repeat_region /note="L2 copies 2 mer aa 72% conserved"
                3251. .3274
repeat_region /note="L2 copies 2 mer gt 95% conserved"
                4009. .4296
repeat_region /note="AluJo repeat: matches 1. .288 of consensus"
                5975. .6130
repeat_region /note="AluJsp/q repeat: matches 179. .312 of consensus"
                7439. .7827
misc_feature /note="match: GSS: Em:AQ037370"
                8681. .8897
repeat_region /note="L1MC1 repeat: matches 3957. .4167 of consensus"
                8898. .9219
repeat_region /note="AluJy repeat: matches 1. .311 of consensus"
                9220. .10052
repeat_region /note="L1MC1 repeat: matches 4167. .5034 of consensus"
                10061. .10139
repeat_region /note="MAD1 repeat: matches 2. .80 of consensus"
                10142. .11371
repeat_region /note="L1MC1 repeat: matches 5086. .6332 of consensus"
                12051. .12391
repeat_region /note="MER2 repeat: matches 1. .345 of consensus"
                12822. .12959
repeat_region /note="L2 repeat: matches 1. .345 of consensus"
                12847. .14238
repeat_region /note="MER69 repeat: matches 2. .1368 of consensus"
                14428. .14575
repeat_region /note="MIR repeat: matches 82. .251 of consensus"
                14895. .15194
repeat_region /note="AluJsq repeat: matches 7. .310 of consensus"
                15398. .15458
repeat_region /note="L2 repeat: matches 2652. .2709 of consensus"
                15688. .15995
repeat_region /note="AluJo repeat: matches 1. .292 of consensus"
                17565. .18421
repeat_region /note="L1PA14 repeat: matches 5258. .6149 of consensus"
                19186. .19499
repeat_region /note="MER63 repeat: matches 727. .1059 of consensus"
                19704. .19809
repeat_region /note="L2 repeat: matches 2585. .2707 of consensus"
                21626. .21778
repeat_region /note="L2 repeat: matches 2593. .2718 of consensus"
                24403. .24569
misc_feature /note="match: GSS: Em:AQ008385"
                25408. .25554
repeat_region /note="MER91A repeat: matches 2. .153 of consensus"
                25931. .26268
misc_feature /note="match: GSS: Em:AQ723618"
                25934. .26472
misc_feature /note="match: GSS: Em:AQ186968"
                26068. .26212
repeat_region /note="match: GSS: Em:AQ633097"
                26145. .26321
repeat_region /note="MIR repeat: matches 77. .262 of consensus"
                26393. .26544
repeat_region /note="MIR repeat: matches 54. .207 of consensus"
                26915. .27138

repeat_region /note="AluJo repeat: matches 84. .308 of consensus"
                27215. .27515
repeat_region /note="AluJsq repeat: matches 14. .305 of consensus"
                27516. .27557
repeat_region /note="L2 copies 2 mer ta 100% conserved"
                27741. .28038
repeat_region /note="AluJo repeat: matches 1. .306 of consensus"
                28216. .28602
repeat_region /note="LTR1J repeat: matches 123. .514 of consensus"
                28977. .29128
repeat_region /note="AluJb repeat: matches 137. .288 of consensus"
                29132. .29221
misc_feature /note="L45 copies 2 mer ag 81% conserved"
                29835. .29921
repeat_region /note="match: GSS: Em:AQ481755"
                30072. .30373
repeat_region /note="L2 repeat: matches 2660. .2745 of consensus"
                31283. .36209
repeat_region /note="AluJo repeat: matches 13. .302 of consensus"
                36210. .36589
repeat_region /note="L1MA2 repeat: matches -41. .4975 of consensus"
                36590. .37914
repeat_region /note="MSTRB repeat: matches 1. .426 of consensus"
                39728. .40083
repeat_region /note="LTR1A2 repeat: matches 24. .372 of consensus"
                40133. .40703
repeat_region /note="MER50 repeat: matches 115. .711 of consensus"
                41144. .41577
misc_feature /note="match: GSS: Em:AQ244745"
                41174. .41811
repeat_region /note="match: GSS: Em:AQ091350"
                41273. .41373
repeat_region /note="HSMAR1 repeat: matches 1. .104 of consensus"
                42032. .42582
repeat_region /note="match: GSS: Em:AQ506183"
                42669. .42753
repeat_region /note="MER87 repeat: matches 1. .539 of consensus"
                42845. .43270
misc_feature /note="match: GSS: Em:AQ586614"
                42890. .43280
misc_feature /note="match: GSS: Em:AQ179179"
                43122. .43319
repeat_region /note="L1MC4 repeat: matches 7892. .7976 of consensus"
                43122. .43319
repeat_region /note="match: GSS: Em:AQ308967"
                44402. .44514
repeat_region /note="L2 repeat: matches 2462. .2583 of consensus"
                44622. .44764
repeat_region /note="MIR repeat: matches 87. .234 of consensus"
                44805. .44931
repeat_region /note="L2 repeat: matches 2609. .2741 of consensus"
                45525. .45525
misc_feature /note="match: GSS: Em:AQ412451"
                45595. .46034
misc_feature /note="match: GSS: Em:AQ514359"
                45602. .45992
misc_feature /note="match: GSS: Em:AQ0801713"
                45724. .46061
repeat_region /note="HAL1 repeat: matches 208. .569 of consensus"
                46151. .46316
repeat_region /note="HAL1 repeat: matches 2. .165 of consensus"
                46524. .47037
misc_feature /note="match: GSS: Em:AQ360084"
                48425. .48510
repeat_region /note="LTR1J repeat: matches 234. .312 of consensus"
                49436. .49584
repeat_region /note="L2 repeat: matches 2535. .2705 of consensus"
```

repeat_region 49668..49871

Query Match 2.1%; Score 23; DB 9; Length 115009;
Best Local Similarity 100.0%; Pred.No. 0.37;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 gcccttgagaagataagagaga 137
|||||
Db 52480 GCCTTGAGAGATTAAGGAGCA 52502

RESULT 11
AC025574
LOCUS Homo sapiens chromosome 12 clone RP11-348M3, WORKING DRAFT
DEFINITION
SEQUENCE, 14 unordered pieces.
AC025574
VERSION AC025574.10 GI:9438393
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K., Bodola,B., Bouck,J., Bowler,S., Brooks,A., Bunay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guerrera,W., Harris,K., Hernandez,J., Hodgson,A., Hogue,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondrjewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucher,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quides,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugan,R., Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wahbah,M., Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155929)
AUTHORS Worley,K.C.
JOURNAL Direct Submission
Submitted (11-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 25, 2000 this sequence version replaced gi:8571540.

COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HALY
Center clone name: RP11-348M3
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-terminator Big Dye 3 of reads
Chemistry: Dye-terminator Big Dye 3 of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 133622 bases at least Q40
Consensus quality: 145733 bases at least Q30
Consensus quality: 150334 bases at least Q20
Estimated insert size: 150694; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 3.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 36972: contig of 36972 bp in length
36973 37072: gap of unknown length
37073 37073: contig of 27829 bp in length
64901 64901: gap of unknown length
65002 65002: contig of 26013 bp in length
91015 91015: gap of unknown length
91115 91115: gap of 14008 bp in length
105123 105123: gap of unknown length
105222 105222: gap of unknown length
113963 113963: contig of 8741 bp in length
114064 114064: gap of unknown length
124894 124894: contig of 10831 bp in length
124895 124895: gap of unknown length
124994 124994: gap of unknown length
132612 132612: contig of 7618 bp in length
132613 132613: gap of unknown length
132713 132713: gap of 6947 bp in length
136659 136659: contig of 6947 bp in length
139759 139759: gap of unknown length
143764 143764: contig of 4005 bp in length
143765 143765: gap of unknown length
143864 143864: gap of unknown length
148062 148062: contig of 4198 bp in length
148063 148063: gap of unknown length
148162 148162: gap of unknown length
151000 151000: contig of 2838 bp in length
151001 151001: gap of unknown length
151101 151101: gap of unknown length
153374 153374: contig of 2274 bp in length
153375 153375: gap of unknown length
153474 153474: gap of unknown length
154810 154810: contig of 1336 bp in length
154811 154811: gap of unknown length
154911 154911: gap of 1019 bp in length.

FEATURES
source 1. 155929
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-348M3"

BASE COUNT 40876 a 36387 c 36205 g 41119 t 1342 others
ORIGIN

Query Match 2.0%; Score 22; DB 2; Length 155929;
Best Local Similarity 100.0%; Pred.No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 899 ctgcccttgatgactatcac 920
|||||
Db 154274 CTGCCCTTGATGACTATCAC 154295

RESULT 12
AL445472/c
LOCUS Human DNA sequence from clone RP13-13705 on chromosome X, complete
DEFINITION
SEQUENCE.
AC025574
VERSION AL445472.14 GI:11544554
KEYWORDS HTG:
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.
JOURNAL Willson,S.
Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: hunquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Dec 4, 2000 this sequence version replaced gi:11493337. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

RP13-13705 is from the library RPC1-13.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PBAC3.6>

IMPORTANT: This sequence is not the entire insert of clone RP13-13705 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-465E19 is at 65250 in this sequence. The true right end of clone RP11-5415 is at 100 in this sequence.

FEATURES

Source

Location/Qualifiers

1..65349

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone="RP13-13705"

/clone_lib="RPC1-13.1"

1..61

/note="Alu repeat: matches 1..61 of consensus"

63..1319

/note="L1PA13 repeat: matches 691..1979 of consensus"

2280..2477

/note="MIR repeat: matches 53..262 of consensus"

2591..4020

/note="L1M3 repeat: matches 4339..5779 of consensus"

4163..4231

/note="MIR2CB repeat: matches 395..462 of consensus"

4466..4594

/note="L2 repeat: matches 2616..2750 of consensus"

4762..5126

/note="L2 repeat: matches 2096..2473 of consensus"

5154..5924

/note="L1M1 repeat: matches 5317..6146 of consensus"

6087..6163

/note="L1P4 repeat: matches 5694..5770 of consensus"

6164..6948

/note="L1M1 repeat: matches 4505..5319 of consensus"

6951..7168

/note="L2 repeat: matches 2477..2691 of consensus"

7270..7487

/note="MIR repeat: matches 28..258 of consensus"

7548..7983

/note="MIR repeat: matches 1..505 of consensus"

8359..8543

/note="MIR repeat: matches 3..178 of consensus"

8788..9100

/note="MIR repeat: matches 3..313 of consensus"

9702..9926

/note="MIR repeat: matches 2..220 of consensus"

10477..10782

/note="MIR repeat: matches 3..309 of consensus"

10919..10962

/note="MIR repeat: matches 3..309 of consensus"

/note="22 copies 2 mer ca 100% conserved"

11429..11564

/note="MIR repeat: matches 1..144 of consensus"

11872..12081

/note="MIR repeat: matches 1..219 of consensus"

12655..14819

/note="L1PA3 repeat: matches 3978..6146 of consensus"

14814..16272

/note="L1PA3 repeat: matches 2500..3962 of consensus"

16861..17345

/note="L2 repeat: matches 1535..2079 of consensus"

17410..17745

/note="L2 repeat: matches 2278..2609 of consensus"

17744..23904

/note="L1PA4 repeat: matches 2..6146 of consensus"

23918..23999

/note="L2 repeat: matches 2571..2688 of consensus"

24679..24792

/note="L1M2 repeat: matches 5690..5796 of consensus"

24834..24984

/note="MIR repeat: matches 9..186 of consensus"

25006..25169

/note="L1M2 repeat: matches 5481..5645 of consensus"

25181..25222

/note="L2 repeat: matches 2652..2694 of consensus"

26036..26135

/note="L2 repeat: matches 2567..2695 of consensus"

27000..27238

/note="L1M4 repeat: matches 4883..5131 of consensus"

27402..31919

/note="L1PA7 repeat: matches 1624..6141 of consensus"

31910..32439

/note="L1MD3 repeat: matches 7149..7711 of consensus"

32466..32595

/note="65 copies 2 mer aa 60% conserved"

32992..33141

/note="MIR20 repeat: matches 62..215 of consensus"

34198..34259

/note="L2 repeat: matches 2682..2737 of consensus"

34260..34430

/note="L1MA2 repeat: matches 6137..6308 of consensus"

34431..34486

/note="L2 repeat: matches 2625..2682 of consensus"

34767..34835

/note="L2 repeat: matches 2629..2699 of consensus"

34869..34906

/note="19 copies 2 mer gt 89% conserved"

35012..35071

/note="L2 repeat: matches 2614..2668 of consensus"

35103..35124

/note="11 copies 2 mer aa 100% conserved"

35136..35454

/note="L1 repeat: matches 1..312 of consensus"

35690..35919

/note="L2 repeat: matches 2537..2749 of consensus"

35920..36216

/note="L1 repeat: matches 1..290 of consensus"

36217..36424

/note="L2 repeat: matches 2324..2537 of consensus"

36495..36590

/note="48 copies 2 mer ac 74% conserved"

36604..36670

/note="MIR repeat: matches 1..69 of consensus"

36671..36752

/note="MIR repeat: matches 1..69 of consensus"

36789..37687

/note="L1PA11 repeat: matches 5274..6165 of consensus"

37666..38802

/note="L1PA10 repeat: matches 5017..6165 of consensus"

38804..39106

/note="L1PA10 repeat: matches 4700..5003 of consensus"

39112..39403

```

repeat_region /note="LIP repeat: matches 5004. .5301 of consensus"
39400. .40335
/note="LIPB repeat: matches -265. .1612 of consensus"
40331. .40862
repeat_region /note="LIPB repeat: matches -943. .411 of consensus"
40863. .41058
repeat_region /note="LIPB repeat: matches -1540. .1344 of consensus"
41071. .41186
repeat_region /note="MTL1-INTERNAL repeat: matches 1098. .1211 of
consensus"
41227. .41350
/note="THE1B-INTERNAL repeat: matches 1518. .1650 of
consensus"
41357. .41824
repeat_region /note="MT1C repeat: matches 6. .464 of consensus"
42031. .42082
repeat_region /note="L2 repeat: matches 2659. .2710 of consensus"
42317. .42553
repeat_region /note="MIR repeat: matches 2. .238 of consensus"
42808. .42869
repeat_region /note="MIR repeat: matches 201. .262 of consensus"
44933. .45093
repeat_region /note="MIR repeat: matches 2. .152 of consensus"
45461. .45702
repeat_region /note="L2 repeat: matches 2246. .2502 of consensus"
45910. .46029
repeat_region /note="L2 repeat: matches 2574. .2710 of consensus"
46560. .47007
repeat_region /note="MT1B repeat: matches 1. .390 of consensus"
47018. .47105
repeat_region /note="L2 repeat: matches 1995. .2095 of consensus"
47284. .47527
repeat_region /note="MIR repeat: matches 3. .254 of consensus"
47649. .47802
repeat_region /note="MIR repeat: matches 65. .215 of consensus"
48031. .48218
repeat_region /note="MERSA repeat: matches 5. .187 of consensus"
48232. .48401

Query Match 1.9%: Score 21; DB 9; Length 65349;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 aaggaagtggaagagcttgag 108
|||||
Db 34593 AAGGAGGTGGAAAGCTTGAG 34573

RESULT 13
AC090521 77644 bp DNA INV 27-FEB-2001
LOCUS Caenorhabditis briggsae cosmid CB002D19, complete sequence.
AC090521
AC090521.1 GI:13129524
VERSION HTG.
KEYWORDS Caenorhabditis briggsae.
SOURCE Caenorhabditis briggsae.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 77644)
AUTHORS Washington University Genome Sequencing Center.
TITLE The C. briggsae Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 77644)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsp1eth@watson.wustl.edu

```

```

FEATURES
SOURCE
1. 77644
/organism="Caenorhabditis briggsae"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="CB002D19"
64453. .64525
/note="codon recognized: AAA"
/product="tRNA-Lys"
BASE COUNT 22191 a 14948 c 1587 g 24918 t
ORIGIN

Query Match 1.9%: Score 21; DB 3; Length 77644;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 451 ctgcacaaacacaaatctca 471
|||||
Db 62656 CTGCCAAAACCAATCTCA 62676

RESULT 14
AC011381 109549 bp DNA HMG 23-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTB-133N3, WORKING DRAFT SEQUENCE.
DEFINITION 7 unordered pieces.
AC011381
AC011381.5 GI:8576058
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HMG; ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 109549)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 109549)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 21, 2000 this sequence version replaced gi:7710549.

COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 115591, H239
Center clone name: CIT9785KB_133N3
-----
Summary Statistics
Consensus quality: 98174 bases at least Q40
Consensus quality: 103469 bases at least Q30
Consensus quality: 105214 bases at least Q20
Estimated insert size: 110000; pulse field gel estimation
Estimated insert size: 108948; sum-of-contrigs estimation
Quality coverage: 5.98 in Q20 bases; pulse field gel estimation
Quality coverage: 6.04 in Q20 bases; sum-of-contrigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
1
* 1231: contig of 1231 bp in length
* 1232 1331: gap of unknown length
* 1332 8134: contig of 6803 bp in length
* 8135 8234: gap of unknown length
* 8235 15598: contig of 7364 bp in length
* 15599 15698: gap of unknown length
* 15699 22852: contig of 7154 bp in length
* 22853 22952: gap of unknown length
* 22953 30470: contig of 7518 bp in length
* 30471 30570: gap of unknown length
* 30571 59501: contig of 28931 bp in length
* 59502 59601: gap of unknown length
* 59602 109549: contig of 49948 bp in length.
Location/Qualifiers
1..109549:
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-133N3"
/clone_lib="CalTech human BAC library B"
BASE COUNT 31123 a 25360 c 23787 g 28626 t 653 others
ORIGIN

```

```

Query Match 1.9%; Score 21; DB 2; Length 109549;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1013 aacatttgataagaagctta 1033
|||||
Db 109083 AACATTGTGATAAGAACTTA 109103

```

```

RESULT 15
AC026800/c AC026800 146206 bp DNA PRI 08-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2377N4, complete sequence.
ACCESSION AC026800
VERSION AC026800.4 GI:14329105
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 146206)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 146206)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 146206)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 146206)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jun 8, 2001 this sequence version replaced gi:13470154.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of sequence;
Estimated Total Number of Errors is 0.1.
STS Content:

```

```

WI-4499 G02896
SHGC-84555 G51780
SHGC-101777 G56263.
Location/Qualifiers
1..146206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2377N4"
BASE COUNT 46774 a 28468 c 27627 g 43337 t
ORIGIN

```

```

Query Match 1.9%; Score 21; DB 9; Length 146206;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 309 tgaacttctggggagcactg 329
|||||
Db 18560 TGAACCTTCTGGGAGGACACTG 18540

```

```

Search completed: February 26, 2002, 13:23:34
Job Time: 9345 sec

```